Numerical Integration and the Constant Strain Condition*

A. E. Frey,[†] C. A. Hall, and T. A. Porsching
Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

Dedicated to Alston S. Householder
on the occasion of his seventy-fifth birthday.

Submitted by G. W. Stewart

ABSTRACT

The finite element method is a very popular method for numerically solving boundary value problems. A sufficient condition for the convergence of this method is the so-called constant strain condition. In this paper we present a proof, using elementary linear algebra, that numerical integration does not affect the ability of an element to satisfy the constant strain condition.

1. INTRODUCTION

Consider the model second order elliptic boundary value problem

$$\begin{aligned} -\nabla\cdot A\nabla u &= f, \qquad (x,y)\in\Omega\subset R^2,\\ u &= 0, \qquad (x,y)\in\partial\Omega, \end{aligned} \tag{1}$$

where A is a 2 by 2 uniformly positive definite symmetric matrix. We assume that f and each a_{ij} is a polynomial of degree at most d_1 in either variable and that Ω is bounded. It is well known [3–5] that the solution u to (1) minimizes the functional

$$F[w]=\frac{1}{2}\int_{\Omega}(\nabla w^T A\nabla w-2fw)\,dx\,dy \tag{2}$$

*This work was supported by the Air Force Office of Scientific Research under Contract #F44620-76-C-0104.

[†]Current address: Department of Mathematical Sciences, Northern Illinois University, DeKalb, Ill. 60115.

0024-3795/80/010151+9$01.75

over the Sobolev space $H_0^1(\Omega)$. A typical application of the finite element method involves decomposing Ω into N subdomains (elements) Ω_i, possibly with curved boundary segments (parametrized as polynomials) to conform to or approximate $\partial\Omega$. A finite dimensional subspace S of $H_0^1(\Omega)$ is then constructed by piecing together functions with domains Ω_i.

A function $w \in S$ when restricted to Ω_i is defined as follows. A one to one mapping

$$\bar{T}_i : \mathbb{S} \equiv [0,1] \times [0,1] \rightarrow \Omega_i$$

is constructed which induces a generalized (r,s) coordinate system on Ω_i:

$$\bar{T}_i(r,s) = \begin{bmatrix} x(r,s) \\ y(r,s) \end{bmatrix}, \qquad 0 \leqslant r,s \leqslant 1. \tag{3}$$

The function w (restricted to Ω_i) is typically chosen to be a polynomial in r and s. Basis polynomials or "shape functions" [5], $\{N_j(r,s)\}_{j=1}^{M}$, are defined on $\mathbb{S}$ by the cardinality condition

$$N_j(r_k, s_k) = \delta_k^j, \qquad 1 \leqslant j,k \leqslant M, \tag{4}$$

where the "nodes" of the element are $\bar{T}_i(r_k, s_k)$, $k = 1,2,\ldots,M$. A function $w \in S$ is defined in Ω_i by

$$w(x,y) = \sum_{j=1}^{M} c_j^i N_j(r,s), \tag{5}$$

where $\bar{T}_i(r,s) = (x,y)$ and $c_j^i = w(\bar{T}_i(r_j, s_j))$. We assume that each N_j and hence w is of degree at most d_2 in either variable.

One then minimizes the functional F over the finite dimensional space S to obtain the coefficient sets $\{\hat{c}_j^i\}_{j=1}^{M}$, $i = 1,\ldots,N$, for the finite element approximation U to u. Since F is a quadratic functional, this minimization leads to a system of P linear equations

$$K\hat{\mathbf{c}} = \mathbf{f}, \tag{6}$$

where P is the total number of nodes in $\Omega \equiv \bigcup_{i=1}^{N} \Omega_i$. The matrix K is "assembled" [5] or constructed from element "stiffness matrices" K^i as

$$[K]_{kl} = \Sigma' [K^i]_{jm}, \qquad 1 \leqslant k,l \leqslant P. \tag{7}$$

A term is present in the sum if and only if there exists a pair of indices (j, m) that label in the local ordering of element Ω_i the identical nodes labeled k and l in the global ordering of Ω. The element stiffness matrix is defined by

$$[K^i]_{jm} = \int_{\mathbb{S}} \nabla N_j^T B^i \nabla N_m \, dr\, ds, \qquad 1 \leqslant j, m \leqslant M, \tag{8}$$

where $B^i = J_i^{-1} A (J_i^{-1})^T \det J_i$ and J_i is the Jacobian matrix associated with $\bar{T}_i$,

$$J_i = \begin{bmatrix} x_r & x_s \\ y_r & y_s \end{bmatrix}.$$

2. THE PERTURBED SYSTEM

A popular choice for $\bar{T}_i(r,s)$ in (3) is the *isoparametric* mapping [3–5],

$$\bar{T}_i(r,s) = \sum_{j=1}^{M} \begin{bmatrix} x_j \\ y_j \end{bmatrix} N_j(r,s), \tag{9}$$

in which case the entries of J are polynomials in r and s. It is obvious that B^i contains *rational* functions of r and s, and herein lies the problem. It is generally impossible to compute the entries of K^i *exactly* using standard quadrature formulae designed to integrate polynomials of a certain degree. Also, the generality of $\bar{T}_i$ in (9) prohibits the construction of any special formulae for the rational integrands in (8), and one uses quadrature formulae (such as Gaussian quadrature [4, 5]) of the generic form

$$\int_{\mathbb{S}} I(r,s)\, dr\, ds \doteq \sum_{k=1}^{K} \omega_k I(r_k, s_k). \tag{10}$$

We note that if Ω_i is rectangular, then J_i is a matrix of constants and the integrand in (8) is a polynomial of degree at most $d \equiv (d_1+2)d_2$ in either variable. Henceforth we assume the quadrature formula in (10) has been chosen so that polynomials of degree d are integrated exactly. For elements Ω_i with curved boundaries such formulae introduce an error matrix E into the system (7), and we designate the solution of the perturbed system by $\tilde{\mathbf{c}}$, i.e.,

$$(K+E)\tilde{\mathbf{c}} = \mathbf{f}. \tag{11}$$

Note that the entries of f are sums of integrals of the form

$$\int_{\Omega_i} f N_j \det J_i \, dr\, ds,$$

which by assumption have polynomial integrands and can be integrated exactly by the numerical quadrature (10).

The system (11) is the system that the computer would solve (ignoring roundoff error) to obtain the "computed" finite element approximation

$$U_c(x,y) = \sum_{j=1}^{M} \tilde{c}_j^i N_j(r,s), \qquad (x,y) \in \Omega_i, \tag{12}$$

$i = 1, 2, \ldots, N$. In contrast, *if* all integrations were done exactly, then the solution to (7) would yield the "true" finite element approximation

$$U(x,y) = \sum_{j=1}^{M} \hat{c}_j^i N_j(r,s), \qquad (x,y) \in \Omega_i. \tag{13}$$

To assess the size of E^i, consider the 8-node element in Fig. 1 and assume $A = I$ in (1). The element shape functions $N_j(r,s)$ are quadratic in r and s. In Table 1 we include some of the entries of the "exact" stiffness matrix K^i (computed using composite integration) as well as the corresponding entries of the stiffness matrix $K^i + E^i$ for 2×2 Gauss and 3×3 Gauss quadratures. We note that $\max_{k,l} |E_{kl}^i| = 0.333$ (10% relative error) in the first case and 0.068 (0.5% relative error) in the second.

8 NODE 2-D ELEMENT
(0.00, 0.30)
(1.00, 0.00)
(0.50, 1.00)
(0.00, 0.00)
(0.50, 0.15)
(1.00, 0.50)
(0.25, 0.00)
(0.00, 0.55)

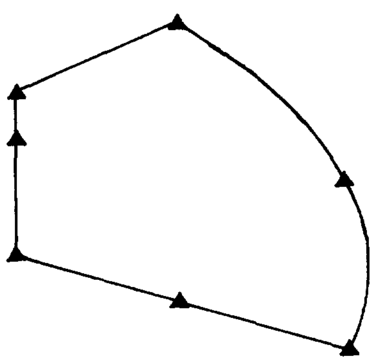

FIG. 1.

TABLE 1

INTEGRATION ERRORS

(j,m)	$[K^i]_{jm}$	$[K^i+E^i]_{jm}$ 2×2 Gauss	$[K^i+E^i]_{jm}$ 3×3 Gauss
(1,1)	1.52545	1.50687	1.51793
(4,4)	1.27413	0.97874	1.20602
(2,3)	0.61383	0.67563	0.61449
(4,6)	−0.43685	−0.56271	−0.44188
(7,7)	3.09015	2.75677	3.08130
(7,8)	−0.01709	−0.00772	−0.00435

3. CONSTANT STRAIN

Historically, the finite element method had its origins in engineering mechanics, and the functionals analogous to F in (2) were integrals whose integrands involved "strains" or rates of change of displacement. For our problem, u_x and u_y play the role of "strain." Very early (e.g. [2]) it was argued that a sufficient condition for convergence of the finite element method was that the element type used must be able to "reproduce constant strain." That is, S contains arbitrary linear polynomials.

It is well known [3, Theorem 8.2] that there exists a constant C such that

$$\|u-U\|_{H^1(\Omega)} \leq C\|u-W\|_{H^1(\Omega)} \tag{14}$$

for all $W\in S$. Further [3, Theorem 6.8], if u is sufficiently smooth, only quasiuniform meshes are considered, and if the constant strain condition is satisfied, then there exists a function $W\in S$ such that

$$\|u-W\|_{H^1(\Omega)} = O(h) \tag{15}$$

as the mesh gauge $h\to 0$. Combining (14) and (15), we see that mathematically the *constant strain condition implies convergence.*

The question of whether errors in computing K^i or K affect an element's ability to reproduce constant strain is answered by a corollary of the following theorem.

THEOREM 1. *Let ϕ be a function with domain Ω and ϕ^i be the $M\times 1$ vector whose kth entry is $\phi(\overline{T}_i(r_k,s_k))$. If the element type being used satisfies the constant strain condition, then $\mathbf{1}^i$, $\mathbf{x}^i$ and $\mathbf{y}^i$ belong to the null space of E^i.*

Proof. Let $\mathbf{c}^i=(c_1^i,c_2^i,\ldots,c_M^i)$ be the vector of coefficients associated with some $w\in S$ [cf. (5)]. Let $P_k\equiv(r_k,s_k)$, $k=1,\ldots,K$, be the integration points used in (10); then

$$\left[(K^i+E^i)\mathbf{c}^i\right]_j=\sum_{m=1}^{M}\left[\sum_{k=1}^{K}\omega_k\left(\nabla N_j^TB^i\nabla N_m\right)(P_k)\right]c_m^i$$

$$=\sum_{k=1}^{K}\omega_k\nabla N_j^T(P_k)B^i(P_k)\left[\sum_{m=1}^{M}c_m^i\nabla N_m(P_k)\right]. \tag{16}$$

But we recognize the expression in square brackets above as $\nabla w(P_k)$.

From (8) we also have

$$\left[K^i\mathbf{c}^i\right]_j=\sum_{m=1}^{M}\left[\int_{\mathbb{S}}\nabla N_j^TB^i\nabla N_m\,dr\,ds\right]c_m^i$$

$$=\int_{\mathbb{S}}\nabla N_j^TB^i\left[\sum_{m=1}^{M}c_m^i\nabla N_m\right]dr\,ds$$

$$=\int_{\mathbb{S}}\nabla N_j^TB^i\nabla w\,dr\,ds. \tag{17}$$

By the constant strain condition, there are constants c_j^i such that in (5), $w(x,y)=1$, $(x,y)\in\Omega_i$. In fact in view of (4), $\mathbf{c}^i=\mathbf{1}^i$. For this choice of w, $\nabla w(x,y)=0$ for all $(x,y)\in\Omega_i$. Hence from (16) and (17),

$$(K^i+E^i)\mathbf{1}^i=0=K\mathbf{1}^i \tag{18}$$

and $\mathbf{1}^i$ belongs to the null space of E^i.

Next consider the function $\phi(x,y)=x$. By the constant strain condition there exist constants c_j^i such that in (5), $w(x,y)=x$, $(x,y)\in\Omega_i$. In fact $\mathbf{c}^i=\mathbf{x}^i$. Then (16) becomes

$$\left[(K^i+E^i)\mathbf{x}^i\right]_j=\sum_{k=1}^{K}\omega_k\nabla N_j^T(P_k)B^i(P_k)\nabla x(P_k) \tag{19}$$

while from (17)

$$\left[K^i\mathbf{x}^i\right]_j=\int\nabla N_j^TB^i\nabla x\,dr\,ds. \tag{20}$$

Now $E^i\mathbf{x}^i=0$ will hold if we can show the integrand in (20) is a polynomial of degree less than or equal to d in either variable.

From the definition of B^i we have

$$B^i=\frac{1}{\det J_i}\begin{bmatrix} a_{11}\,y_s^2-2a_{12}x_sy_s+a_{22}x_s^2 & \text{symmetric} \\ \begin{matrix}-a_{11}\,y_ry_s+a_{12}x_ry_s \\ +a_{12}x_sy_r-a_{22}x_rx_s\end{matrix} & a_{11}\,y_r^2-2a_{12}x_ry_r+a_{22}x_r^2 \end{bmatrix}$$

and

$$B^i\nabla x=\frac{1}{\det J_i}\begin{bmatrix}\det J_i(a_{11}\,y_s-a_{12}x_s) \\ \det J_i(a_{12}x_r-a_{11}\,y_r)\end{bmatrix}=\begin{bmatrix}a_{11}\,y_s-a_{12}x_s \\ a_{12}x_r-a_{11}\,y_r\end{bmatrix}. \tag{21}$$

From the assumptions on A and the form of $\overline{T}_i$, the entries of $B^i\nabla x$ as well as the integrand in (20) are polynomials of degree less than d in both r and s. Hence the numerical quadrature in (10) applied to (20), which yields (19), is exact and we have

$$(K^i+E^i)\mathbf{x}^i=K^i\mathbf{x}^i. \tag{22}$$

Thus $\mathbf{x}^i$ belongs to the null space of E^i.

The argument for $\phi(x,y)=y$ is similar and is omitted. ■

Now if the constant strain condition is satisfied, then any linear function u has a representation in S, and it follows from (14) that the finite element approximation U produced by exact integrations is identical to u. As a corollary to the above Theorem we also have

COROLLARY. *Suppose that the true solution to* (1) *is some linear function of x and y, say*

$$u(x,y)=a+bx+cy. \tag{23}$$

If the element type being used satisfies the constant strain condition, then the computed finite element approximation U_c is also linear and $U_c=u$.

Proof. If u is linear, then the true finite element solution

$$U(x,y)=\sum_{j=1}^{M}\hat{c}_j^iN_j(r,s), \qquad (x,y)\in\Omega_i,$$

is linear and $\hat{\mathbf{c}}^i=a\mathbf{1}^i+b\mathbf{x}^i+c\mathbf{y}^i$.

Numerical integration leads to the system (11) where the error matrix E satisfies [cf. (7)]

$$[E]_{kl} = \Sigma'[E^i]_{jm}, \qquad 1 \leqslant k, l \leqslant P. \tag{24}$$

The solution $\hat{\mathbf{c}}$ to the unperturbed system (6) also is a solution of (11) in this case, since

$$[E\hat{\mathbf{c}}]_k = \sum_{l=1}^{P} [E]_{kl}\hat{c}_l = \Sigma' \sum_{m=1}^{M} [E^i]_{jm}\hat{c}_m^i = 0$$

by the above theorem. Hence $\tilde{\mathbf{c}} = \hat{\mathbf{c}}$ and $U_c = U = u$. ■

We recall that the numerical quadrature in (10) was chosen to integrate exactly polynomials of degree d in either r or s, so that for rectangular elements the stiffness matrix K^i would be computed exactly (up to roundoff). On the other hand, close scrutiny of (20) shows that the integrand is in fact a polynomial of degree at most $d-1$ in either r or s. Therefore we conclude that Theorem 1 remains valid even if the numerical quadrature in (10) is chosen so as to integrate polynomials of at most degree $d-1$ in either variable. The entries of $\mathbf{f}$ may involve polynomial integrands of degree d and should be handled as before. For example, if Ω_i is a square and $A = I$, then 2×2 Gauss quadrature produces a nontrivial error matrix E^i for the serendipity element [5], since $d = 4$. However, constant strain is preserved.

The above theorem and corollary extend straightforwardly to other functionals which contain only first order derivatives (e.g. plane stress and plane strain [5, Chapter 4]) and to three dimensional domains [1].

If higher order elements are used in which some degrees of freedom are other than the value of the function at a node (e.g. w_x, w_y, w_{xy} in the Hermite family of elements [3, §6.6]), then the theorem is still valid because the vectors $\mathbf{1}^i$, $\mathbf{x}^i$ and $\mathbf{y}^i$ can be replaced by the vectors $\mathbf{c}^i$ determined so that in (5) $w(x, y) = 1$, x and y respectively. The Corollary again establishes that constant strain is preserved under numerical integration for such elements.

A more detailed discussion of the constant strain condition for fourth order as well as second order problems is given in [1].

REFERENCES

1 A. E. Frey, A C^1 curved plate element which satisfies constant strain, Ph.D. Dissertation, Univ. of Pittsburgh, 1978.

2 G. P. Bazeley, Y. K. Cheung, B. M. Irons and O. C. Zienkiewicz, Triangular elements in bending-conforming and non-conforming solutions, in *Proceedings of*

the Conference on Matrix Methods in Structural Mechanics, Air Force Inst. Tech., Wright Patterson A. F. B., Ohio, 1965.
3 J. T. Oden and J. N. Reddy, *Mathematical Theory of Finite Elements*, Wiley, New York, 1976.
4 G. Strang and G. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, 1973.
5 O. C. Zienkiewicz, *The Finite Element Method*, McGraw-Hill, London, 1977.

Received 1 July 1978; revised 25 May 1979